U.S. ATLAS Northeast Tier 2 Renewal Proposal

The bulk of the U.S. computing resources for the ATLAS LHC experiment are organized into a single Tier 1 center and five Tier 2 centers including the Northeast Tier 2 center (NET2) located in Boston, Massachusetts. Over the past four years, physics and IT groups at Boston University and Harvard University have collaborated to build up and operate the NET2 in close collaboration with our sister Tier 2 centers and with the U.S. ATLAS Tier 1 center at Brookhaven National Laboratory. Building on this successful work and on continuing active involvement in ATLAS by both Boston University and Harvard, we propose to continue our work through 2016, maintaining continuous operation and expanding our facilities in accordance with ATLAS planning.

Current and Future Facilities

The original NET2 site and current Boston University component of NET2 consists of 700 cores of IBM blades, 1.1 PB of GPFS storage, a Cisco 6509 router, a 16 core 64GB interactive node and a variety of additional servers. This equipment is currently in the Physics Research Building at 3 Cummington Street on the BU campus. The Harvard NET2 gatekeeper and server node and the Harvard Tier 3 are located at 60 Oxford Street on the Harvard campus. As of our most recent hardware additions, the majority of our current computing capacity (1024 Nehalem cores) is located in a Harvard rented facility at 1 Summer Street in downtown Boston and is managed as part of the Harvard Odyssey cluster. Wide area networking for NET2 is provided by a dedicated 10 Gb/s fiber between our own Cisco 6509 router and the NoX Metro Ring (a 128 fiber ring connecting Boston University, Harvard and MIT). Networking between the NET2 storage elements at BU and the Harvard worker nodes is carried on the two 10Gb/s fibers connecting Odyssey with 1 Summer Street internet hub (1 Summer street is one of the major internet hubs in the northeastern U.S.). The bandwidth between the BU GPFS storage and the Harvard worker nodes has been measured at greater than 800 MB/s disk-to-disk. Effectively the networking between our BU and HU sites is as good as in-cluster local networking. With 500 additional cores contributed from the Harvard Odyssey cluster, NET2 can currently run approximately 2700 simultaneous Panda jobs and has 1.1 PB of storage in GPFS. One of the main milestones of 2010 was showing that facility could function at full capacity even under heavy load from data transfers from BNL to NET2. Images and videos of the NET2 sites are available at http://atlas.bu.edu/NET2-tour.

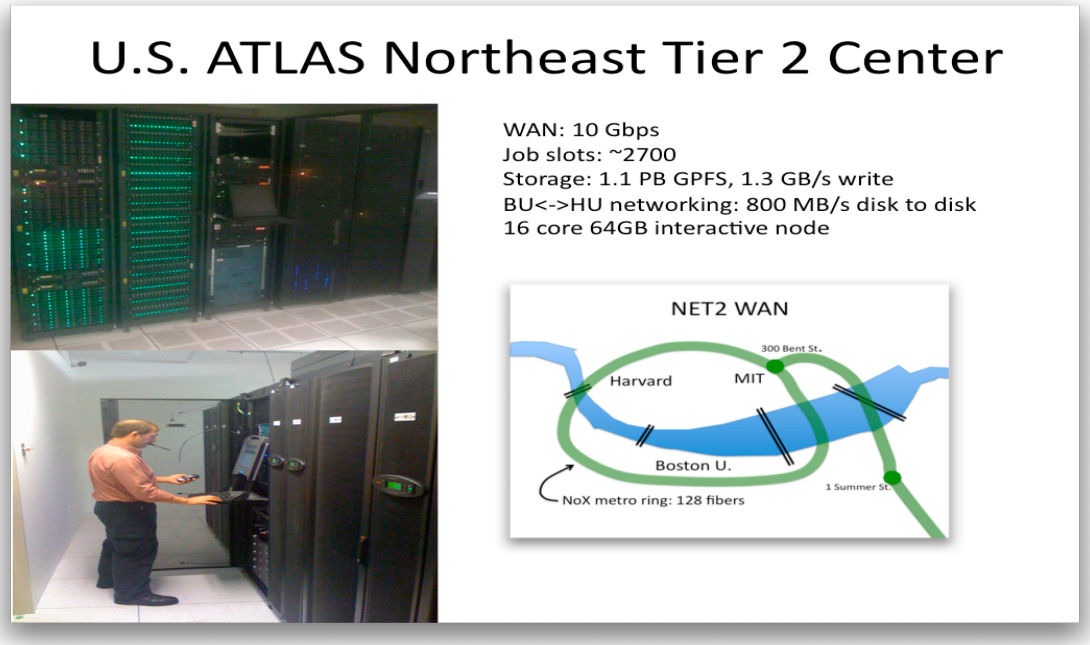

Figure 1. Photos of the the Boston University NET2 machine room (upper left), part of the Harvard Odyssey cluster at 1 Summer Street (lower left) and a schematic diagram of the NET2 wide area networking showing neighboring Universities, internet hubs and the Charles river in blue.

As of October 2010, construction began on a large new, joint Boston University/Harvard/MIT/Northeastern/UMASS HPC facility in Holyoke Massachusetts. When completed in 2012, the facility will be capable of powering and cooling 648 racks with an average power of 14 kW per rack using hydroelectric and nuclear power only. In addition to being carbon-neutral, the new facility will reduce the electricity costs to the Universities by more than a factor of two. The NET2 teams have had a major role in the planning for this facility and will continue to have a role in governing the facility over time. Professors Shank, Rebbi, Brower (all of the BU Center for Computational Sciene) and Professor Bresnahan (director of SCV) are on the Holyoke green HPCC center governing committee. We intend to move the bulk of both our current BU and HU computing to Holyoke beginning late in 2012.

Collaboration

Planning, operations, monitoring and problem response for the two NET2 sites is done in collaboration with two highly experienced IT groups: the Scientific Computing and Visualization group (SCV: http://www.bu.edu/tech/research/scv/) at Boston University and the Faculty of Arts and Sciences Research computing group (FAS-IT: http://rc.fas.harvard.edu/) at Harvard. Each of these groups provides a primary contact who spends most of their time on ATLAS and who can call upon colleagues for temporary increases in activity for planning, for installation and commissioning new equipment and for providing rapid response to operational problems. Both the SCV and FAS-IT groups have years of experience dealing with a variety of hardware and have

proven track records in large scale computing. For example, just Odyssey cluster managed by FAS-IT has roughly the same capacity (currently 12,000 cores) as is typically used in all of U.S. Panda production and runs typically 50,000 batch jobs per day. In total, our experience has been that approximately 3.1 FTE of effort is required to keep NET2 operational, coordinated with the other Tiers and growing over time. Project funds pay for 1.5 FTE of this total. The following table lists the people involved in their project, their role and approximate time spent on the project.

Boston University:

Person	Title	Role	FTE
Jim Shank	Research Professor	Principle Investigator	0.05
Saul Youssef	Research Associate Professor	NET2 management and operations	1.00
Augustine Abaris	Senior Systems Administrator	BU systems administration	0.75
Russ Wolf	Senior Systems Administrator	Rapid response, upgrades	0.10
Wayne Gilmore	Senior Systems Administrator	HVAC, upgrades	0.10
Mike Dugan	Manager, Systems Administrator	Rapid response, planning, upgrades	0.10
Eric Gauthier	Engineering and Operations Manager	BU networking for NET2	0.05

Harvard University:

Person	Title	Role	FTE
John Huth	Professor	Principle Investigator	0.05
James Cuff	Director of Research Computing and Chief Technology Architect	Director, FAS-IT research computing	0.07
Suvendra Dutta	Director of Research Computing in the Physical Sciences	Management of HU Tier 2/3 facilities	0.07
John Brunelle	Software Engineer	HU Tier 2/3 operations	0.80
Chris Walker	Research Computing Associate	Odyssey management and operations	0.07

University Contributed Resources

Both Boston University and Harvard University have contributed quite substantially to the costs of building, maintaining and operating NET2 both in terms of paying salaries, providing hardware and absorbing hardware-related costs.

Boston University hardware-related contributions

Power for IT and AC equipment	**$ 80,000** per year
Networking fee for NoX 10Gpbs fiber	**$ 6,500** per year
HVAC maintenance contract	**$ 12,000** per year
Infrastructure fund	**$ 180,000**
Moving costs to Holyoke	**$ TBD**

Harvard University hardware-related contributions

Power for IT and AC equipment	**$ 90,000** per year
Networking fees	**$ 6,000** per year
HVAC maintenance	**$ 5,000** per year
Two APC in-row cooling units	**$ 20,000**
Two racks for 8 Nehalem Blade chassis	**$ 8,000**
Additional port for 10Gbps WAN	**$ 5,000**
Rent at 1 Summer street facility	**$ TBD**
Moving costs to Holyoke	**$ TBD**

In addition to hardware-related costs, Boston University provides for the salaries of an estimated additional 0.6 FTE and Harvard provides for the salaries of an additional 1.0 FTE. Since 2009, Harvard has also provided NET2 with 500 additional cores from the Odyssey cluster for PandDA production and/or analysis jobs. The director of Harvard FAS research IT has indicated that FAS-IT will continue to provide opportunistic resources to the project at the level of approximately 4% of the full Odyssey cluster as it grows over time through the period 2012-2016. Odyssey is currently 12,000 cores and is expected to grow at a rate of approximately 2000 cores per year through 2016.

History and Projected Resources

The following table shows the history of NET2 funding and delivered computing capacity and useable storage to ATLAS. Listed are total project funding, FTE funded by the project and total FTE working, dollars spend on worker nodes, dollars spend on storage and the resulting computing capacity (in HS06 units) and useable Terabytes.

	Project $	**FTE**	**HS06 $**	**TB $**	**HS06**	**TB**
2007-2008	$600,000	1.5/2.0	$ 56,000	$130,000	3,000	186
2008-2009	$600,000	1.5/2.5	$ 96,000	$186,000	7,900	475
2009-2010	$600,000	1.5/3.1	$290,000	$ 96,000	23,900	1,100
2010-2011	$600,000	1.5/3.1	$105,000	$157,000	30,400	1,730
2011-2012	$609,000	1.5/3.1	$105,000	$157,000	46,000	2,420
2012-2013	$618,000	1.5/3.1	$105,000	$157,000	43,700	2,240
2013-2014	$627,000	1.5/3.1	$105,000	$157,000	43,700	3,390
2014-2015	$637,000	1.5/3.1	$105,000	$157,000	52,200	4,240
2015-2016	$646,000	1.5/3.1	$105,000	$157,000	63,800	5,350

For future years, we have assumed a 1.5% annual budget increase to be applied to salaries, a 60/40 spending ratio on storage/worker nodes, an assumed three year doubling

time for both HS06 and useable TB per dollar and retirement of both storage and worker nodes after three years of operation.

Throughout our first four years, we have maintained nearly 100% operational availability of our site and consistently made all of the above resources available to U.S. and international ATLAS via the integrated PanDA and DQ2 systems. NET2 consistently makes a significant contribution to U.S. ATLAS computing.

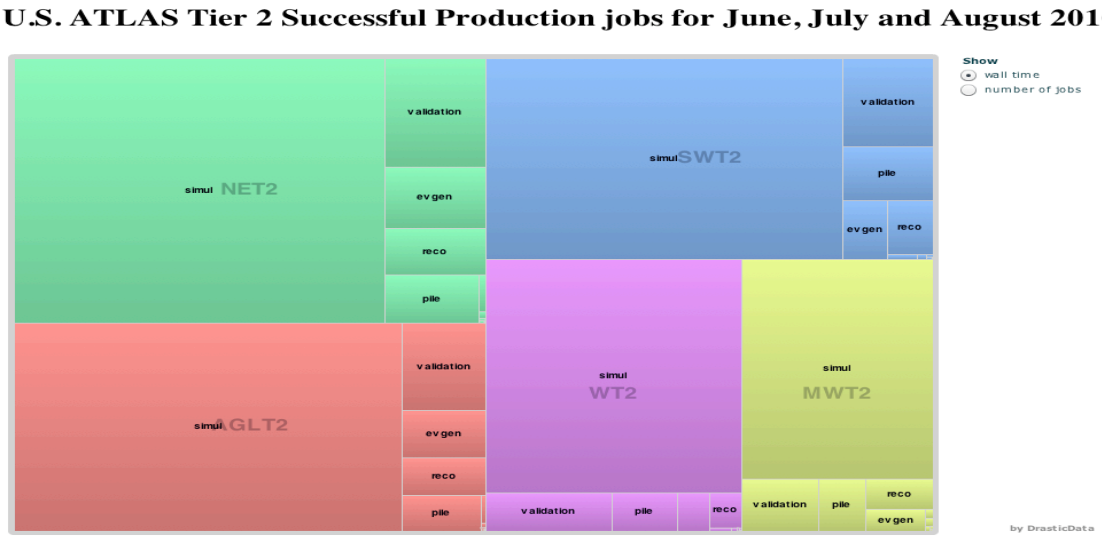

Figure 1. Successful production jobs at the US Tier 2s in the three months after the Harvard Nehalem racks were put into production.j

Figure 1, for example, shows U.S. ATLAS Tier 2 PanDA production in the three months following the turn on of the two Nehalem racks at Harvard.

Infrastructure Software Development

In addition to our other activities, we have undertaken two major infrastructure related software projects during the past years. Pacman is a software installer which is used by ATLAS, OSG and Teragrid among others (http://physics.bu.edu/pacman). There have been many millions of downloads of Pacman to more than 50 countries since it's original introduction in 2001. Development in the past few years consisted of introducing a system of mirrors used providing fast local installation of ATLAS releases and the introduction of self-contained self-installing *pacballs* which allows software environments to be captured in a single file and moved within ATLAS using the DDM system like any other data. Both Pacman mirrors and pacballs have been adopted by ATLAS have been used in essentially flawless operation since their introduction. Pacman has been completely stable for the past two years and work has concentrated on the *Egg software project* (http://egg.bu.edu). Egg is based on an insight into distributed computing originally developed in collaboration with computer scientists at Harvard and with funding from an NSF ITR. In the context of ATLAS and NET2, Egg provides a

simple coherent interface to all parts of the infrastructure and has actively been used to help manage NET2 while we were developing the software. With the 1.0 release imminent, Egg is already beginning to capture wider interest in ATLAS as a potentially powerful way of providing a simple, coherent interface to the whole ATLAS infrastructure for monitoring and other purposes. The same potential for success applies to many situations beyond ATLAS as well.

Interactive user support and Tier 3 activities

Although supporting local users is not necessarily part of the Tier 2 mandate, we have always supported a small community of ATLAS collaborators who work on muon system commissioning and alignment as well as on a variety of related physics topics. These colleagues have made small scale informal use of our Tier 2 batch system directly and occasional heavy use of our Panda analysis sites. In the past year, our colleagues have also received funding for an ATLAS Tier 3 at Harvard and for both an ATLAS and a CMS Tier 3 at Boston University. In the BU case, Tier 3 hardware funds will be used to add NET2 worker nodes and corresponding PBS queues with local user priority. Both the Harvard and Boston Tier 3s will be managed by the same NET2 experienced teams described here, thus building Tier 3 facilities upon our existing infrastructure in an cost-effective way.